

# Comparing galaxy clustering in Horizon-AGN simulated light-cone mocks and VIDEO observations

P. W. Hatfield<sup>1</sup>,<sup>1</sup>★ C. Laigle<sup>2</sup> M. J. Jarvis<sup>1</sup>,<sup>2,3</sup> J. Devriendt<sup>2</sup> I. Davidzon<sup>4</sup> O. Ilbert<sup>5</sup>  
C. Pichon<sup>6,7,8</sup> and Y. Dubois<sup>6</sup>

<sup>1</sup>Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK

<sup>2</sup>Department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

<sup>3</sup>Department of Physics, University of the Western Cape, Bellville 7535, South Africa

<sup>4</sup>IPAC, California Institute of Technology, Mail Code 314-6, 1200 East California Boulevard, Pasadena, CA 91125, USA

<sup>5</sup>Laboratoire d'Astrophysique de Marseille (LAM), CNRS, Aix-Marseille University, F-13388, Marseille, France

<sup>6</sup>Institut d'Astrophysique de Paris, CNRS, UMR 7095, Sorbonne Universités, 98 bis bd Arago, F-75014 Paris, France

<sup>7</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

<sup>8</sup>Korea Institute for Advanced Study (KIAS), 85 Hoegiro, Dongdaemun-gu, Seoul 02455, Republic of Korea

Accepted 2019 October 7. Received 2019 September 9; in original form 2019 April 12

## ABSTRACT

Hydrodynamical cosmological simulations have recently made great advances in reproducing galaxy mass assembly over cosmic time – as often quantified from the comparison of their predicted stellar mass functions to observed stellar mass functions from data. In this paper, we compare the *clustering* of galaxies from the hydrodynamical cosmological simulated light-cone Horizon-AGN to clustering measurements from the VIDEO survey observations. Using mocks built from a VIDEO-like photometry, we first explore the bias introduced into clustering measurements by using stellar masses and redshifts derived from spectral energy distribution fitting, rather than the intrinsic values. The propagation of redshift and mass statistical and systematic uncertainties in the clustering measurements causes us to underestimate the clustering amplitude. We then find that clustering and halo occupation distribution (HOD) modelling results are qualitatively similar in Horizon-AGN and VIDEO. However, at low stellar masses, Horizon-AGN underestimates the observed clustering by up to a factor of  $\sim 3$ , reflecting the known excess stellar mass to halo mass ratio for Horizon-AGN low-mass haloes, already discussed in previous works. This reinforces the need for stronger regulation of star formation in low-mass haloes in the simulation. Finally, the comparison of the stellar mass to halo mass ratio in the simulated catalogue, inferred from angular clustering, to that directly measured from the simulation validates HOD modelling of clustering as a probe of the galaxy–halo connection.

**Key words:** techniques: photometric – galaxies: evolution – galaxies:haloes – galaxies: star formation.

## 1 INTRODUCTION

Parallel to the success of wide-field surveys in *observing* large numbers of galaxies, there have also been great advances in *simulating* large numbers of galaxies in cosmological simulations (see Somerville & Davé 2015, for a review). Cosmological simulations typically model a fraction of the Universe as a finite cube with periodic boundary conditions, take some set of initial conditions and some formulation of the physics of the Universe that the simulator is interested in capturing, and let this realization of the universe evolve

from the beginning of time to  $z = 0$ . Critically, simulations are the only way to solve the full growth of cosmological structure into the strongly non-linear regime, including baryonic physics (down to the resolution limit of the simulation), the analytic derivation being possible only for dark matter (DM) in the linear or weakly non-linear regime (see e.g. Peebles 1980). Cosmological simulations have dramatically changed over the last  $\sim 50$  yr due to the improvement of computing power, from being only able to qualitatively capture the large-scale structure of the Universe to being able to capture in great detail a wide range of baryonic physics processes.

Galaxy populations can be either ‘grafted’ on a DM  $N$ -body simulation according to some prescriptions (e.g. GALFORM, Bower et al. 2006; SAGE, Croton et al. 2006, 2016; SESAM, Neistein &

\* E-mail: [peter.hatfield@physics.ox.ac.uk](mailto:peter.hatfield@physics.ox.ac.uk)

Weinmann 2010; the model of Guo, Zehavi & Zheng (2012) in the Millennium Simulation, Springel et al. 2005) or simulated directly by following the physics of gas and star formation in hydrodynamical simulations (e.g. Horizon-AGN, Dubois et al. 2014; EAGLE, Schaye et al. 2015; McAlpine et al. 2016; ILLUSTRIS, Vogelsberger et al. 2014; MUFASA, Davé, Thompson & Hopkins 2016; MASSIVEBLACK-II, Khandai et al. 2015). The first method, called semi-analytic models (SAMs), provides the opportunity to populate very large volumes at low computational expense, but relies on a large number of tuneable parameters, as the rich diversity of galaxy properties has to be inferred from the DM halo mass and merger trees only. Hydrodynamic simulations, in contrast, seek to incorporate physics in a more fundamental way down to the smallest possible scale by consistently following simultaneously DM, gas, and stars, but are therefore vastly more computationally expensive, and are limited to smaller volumes (typically several hundreds of Mpc for the largest ones).

No simulation can capture all physics present in the real Universe. Instead, we wish to investigate how successful a simulation is in recreating observations and understand which aspects of galaxy formation physics are correctly reproduced and which aspects are currently missing. One approach is to use one-point statistics as a measure of how successful a simulation is at recreating the observed stellar mass function (see e.g. Schaye et al. 2015; Kaviraj et al. 2017), star formation rates (e.g. Katsianis, Tescari & Wythe 2016), morphology (e.g. Dubois et al. 2016), and scaling relations (e.g. Jakobs et al. 2018). One-point statistics are, however, not sufficient to constrain all aspects of galaxy mass assembly. The spatial distribution of the galaxies, as quantified from two-point statistics, gives extra information on the scale at which baryonic processes are important (e.g. Chisari et al. 2018), how galaxies interact with each other and their environment (see e.g. Li et al. 2012, for the comparison of spectroscopic observations and SAMs), and the galaxy–halo connection (Wechsler & Tinker 2018). Comparing the clustering of galaxies in observations to galaxy clustering in hydrodynamic simulations therefore provides additional constraints on galaxy evolution models (see e.g. Saghiha et al. 2017) – this is the focus of this work.

Most previous comparisons between clustering in observations and in simulations have focused on galaxy bias (e.g. Blanton et al. 1999, 2000; Cen & Ostriker 2000; Yoshikawa et al. 2001; Weinberg et al. 2004), the estimation of which is pivotal for making cosmological inferences from these measurements. From the point of view of galaxy evolution, however, the bias does not include all information about the spatial distribution of galaxies especially at non-linear scales. In order to constrain the galaxy evolution model, especially baryonic feedback, we need to probe these small scales, which matter the most for galaxy formation and evolution. These small scales are often modelled with a halo occupation distribution (HOD) phenomenology. Small-scale clustering observations (and corresponding HOD inferences) have been compared to simulation predictions of clustering to test galaxy evolution models in the literature, e.g. McCracken et al. (2007) compare clustering in observations and mock catalogues from SAMs down to scales of  $\sim 0.1$  Mpc in the COSMOS field. Farrow et al. (2015) more recently compared clustering in a  $\sim 180$  deg<sup>2</sup> survey to clustering predictions from GALFORM down to sub-Mpc scales at  $z < 0.5$ . They found that GALFORM correctly predicts the well-known trend of more massive galaxies being more strongly clustered. However, the model made incorrect predictions for other aspects of clustering, in particular the correlation function on the smallest scales, and the clustering of the most luminous galaxies. The impact of observational biases

on clustering measurements and the validity of HOD modelling (how accurate the inferences made from the approach are) have also been examined in the literature. Crocce et al. (2016) compare how inferences from clustering differ when different photometric redshift calculation approaches are used (machine learning based versus template based). Beltz-Mohrmann, Berlind & Sześciw (2019) test the validity of HOD inferences by modelling the clustering in hydrodynamical simulations, and then applying the resulting occupation model to a DM-only simulation. They found that the different halo mass functions (due to the presence of baryons) between the hydrodynamical simulations and the DM-only simulation caused some differences in the resulting clustering, but that small corrections could largely account for this effect (at least for their high-luminosity sample). Finally, Zentner, Hearin & van den Bosch (2014) (see also Hearin, Behroozi & van den Bosch 2016) present concerns that ‘assembly bias’ (the phenomenon that the clustering of DM haloes depends on assembly time as well as mass) could seriously bias inferences from HOD modelling (which typically only includes dependences on halo mass). They suggest that halo mass estimates could be biased by up to  $\sim 0.5$  dex if assembly bias is not incorporated, more than conventional uncertainty calculations would suggest – a serious concern.

In this work, we seek to build on the previous work comparing observations and hydrodynamical simulations (see e.g. Artale et al. 2017 in EAGLE at  $z \sim 0.1$ , Springel et al. 2018 in ILLUSTRIS), by comparing clustering and HOD measurements (as a function of stellar mass) between observations at  $z > 0.5$  and hydrodynamic cosmological simulations. On the simulation side, we work with the Horizon-AGN simulation, a hydrodynamic simulation with gas cooling, star formation, and stellar and active galactic nucleus (AGN) feedback prescriptions (Dubois et al. 2014). The focus is on the comparison of clustering measurements and HOD modelling between VIDEO observations and mock catalogues built from the Horizon-AGN light-cone (Laigle et al. 2019).

Our goal is threefold. First, by using mock catalogues built from the simulation, both including and not including the observational systematics,<sup>1</sup> we can quantify how the propagation of these uncertainties biases the clustering measurements. Secondly, the comparison of galaxy clustering in the real and mock VIDEO observations will allow us to constrain the galaxy evolution model implemented in the simulation. Finally, the measurement of the galaxy–halo connection, derived from the galaxy clustering in the mocks, compared to the connection measured directly in the simulation will help us to confirm that the HOD model is a good way to infer halo mass.

This paper is organized as follows: We describe the observational data we compare the simulation to in Section 2. In Section 3, we briefly describe the Horizon-AGN simulation and the mock catalogue constructed from it. We then measure the correlation function in the simulation mock catalogue for sub-samples corresponding to those considered in VIDEO in Hatfield et al. (2016) in Section 5. We fit HOD models to these simulated observations, and compare and contrast the Horizon-AGN results to those from the VIDEO observations (Section 6). Finally, we discuss the implications of our results for understanding which aspects of galaxy formation physics simulations capture well, and which they do not, and for understanding how confident we can be in inferences from HOD modelling.

<sup>1</sup> Photometric uncertainties and systematics arising when deriving the masses and redshifts from SED fitting.

Unless specified otherwise, we call ‘intrinsic’ the quantities (redshifts and masses) that are directly measured from the simulation, and ‘photometric’ the quantities from the simulation that are derived through spectral energy distribution (SED) fitting on the mock photometry.

## 2 OBSERVATIONS

The VIDEO survey (Jarvis et al. 2013) on the VIRCAM camera on the VISTA telescope (Dalton et al. 2006) is a deep wide-field survey covering three fields in the Southern hemisphere, each carefully chosen for availability of multiband data, totalling  $12 \text{ deg}^2$ . VIDEO is sensitive to similar volumes, redshifts, and stellar masses as Horizon-AGN, permitting a meaningful comparison between the two. The  $5\sigma$  depths of these VIDEO observations (used in Hatfield et al. 2016) in the five bands are  $Z = 25.7$ ,  $Y = 24.5$ ,  $J = 24.4$ ,  $H = 24.1$ , and  $K_s = 23.8$  for a 2 arcsec diameter aperture.

Our catalogue was constructed by combining the VIDEO data set with data from the T0006 release of the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS) D1 tile (Ilbert et al. 2006; Gwyn 2012), which provides photometry with  $5\sigma$  depths (2 arcsec apertures) of  $u^* = 27.4$ ,  $g' = 27.9$ ,  $r' = 27.6$ ,  $i' = 27.4$ , and  $z' = 26.1$  over  $1 \text{ deg}^2$  of the VIDEO XMM3 tile. This data set has already been used in many extragalactic studies to date (e.g. Johnston et al. 2015; White et al. 2015; Hatfield et al. 2016).

The same data processing and sub-sample selection as Hatfield et al. (2016) are used here – see Hatfield et al. (2016) and Jarvis et al. (2013) for a full description of the data set, samples, detection images, detection thresholds, and the construction of the SED templates used, but we briefly summarize the reduction process here. The sources were identified using SEXTRACTOR (Bertin & Arnouts 1996) source extraction software, with 2 arcsec apertures. Photometric redshifts and stellar masses used in this work were estimated using LEPHARE (Arnouts et al. 1999; Ilbert et al. 2006), which fits SED templates to the photometry.

SEXTRACTOR identified 481 685 sources in the CFHTLS-D1  $1 \text{ deg}^2$  with detections in at least one band. As described in Hatfield et al. (2016), we remove sources in regions affected by excess noise and bright stars with a mask. We only include sources with  $K_s < 23.5$ , and make a colour cut around a stellar locus, following the approach of Baldry et al. (2010), to remove stars. VIDEO has a 90 per cent completeness at this depth (Jarvis et al. 2013). McAlpine et al. (2012) estimate that this colour cut leaves stars contributing  $< 5$  per cent of the sample. The final galaxy sample  $\mathcal{C}_{\text{VIDEO}}$  comprises 97 052 sources. The clustering properties of galaxies in this VIDEO-CFHTLS data set have been explored in Hatfield et al. (2016) (where the observed clustering measurements and HOD fits used in this work are taken from) and Hatfield & Jarvis (2017).

## 3 THE VIRTUAL VIDEO-LIKE CATALOGUE

### 3.1 The Horizon-AGN simulation

The full specifications of Horizon-AGN and the details of the physics it incorporates can be found in Dubois et al. (2014, 2016), Kaviraj et al. (2017), and Laigle et al. (2019); we give only a brief description here.

Horizon-AGN was run with the RAMSES adaptive mesh refinement code (Teyssier 2002). It simulates a box with periodic boundary conditions of comoving width  $100 \text{ Mpc } h^{-1}$ , containing  $1024^3$  DM particles (mass resolution of  $8 \times 10^7 M_\odot$ ), compatible with a WMAP7 cosmology (Komatsu et al. 2011). Gas dynamics,

cooling, and heating are followed on the adaptive mesh with a minimum cell size of 1 kpc (constant in physical length). The simulation also follows star formation (star particle mass resolution of  $2 \times 10^6 M_\odot$ ), feedback from stars (stellar winds, Type II and Type Ia supernovae), the evolution of six chemical species, and feedback from AGNs.<sup>2</sup>

AGN feedback can be either ‘radio mode’ or ‘quasar mode’ (Dubois et al. 2012). Radio mode operates at low black hole accretion rates ( $\dot{M}_{\text{BH}}/\dot{M}_{\text{Edd}} < 0.01$ , where  $\dot{M}_{\text{BH}}$  is the black hole accretion rate and  $\dot{M}_{\text{Edd}}$  is the Eddington accretion rate; see Castelló-Mor, Netzer & Kaspi 2016) and injects energy into the inter-galactic medium through bipolar jets. Quasar mode operates at higher accretion rates ( $\dot{M}_{\text{BH}}/\dot{M}_{\text{Edd}} > 0.01$ ) and injects energy isotropically.

### 3.2 Extraction of galaxies and haloes

A light-cone (a simulated box for which one direction is redshift) with opening angle of  $1 \text{ deg}^2$  was constructed on the fly from the simulation (see Pichon et al. 2010), with very fine redshift steps (about 22 000 steps up to  $z = 6$ ). Such light-cone mimics therefore the geometry of observational surveys.

Galaxies have been initially identified on these light-cone slices (about 4000 slices up to  $z = 4$ ) from the distribution of star particles using the ADAPTAHOP structure finder code (Aubert, Pichon & Colombi 2004; Tweed et al. 2009). Structures are selected with a density threshold of 178 times the average matter density and are required to contain at least 50 stellar particles.

DM haloes have been extracted independently on the same slices from the distribution of DM particles, with a density threshold of 80 times the average matter density, and are required to contain at least 100 stellar particles. Galaxies are matched with their closest haloes. To this purpose, the centre of the haloes is accurately defined from the shrinking sphere method (Power et al. 2003).

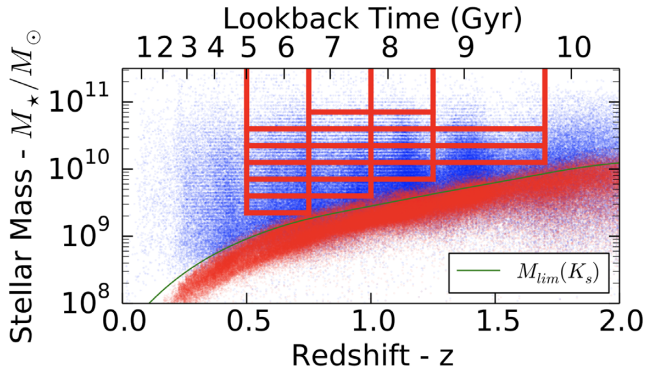
### 3.3 Production of the photometric catalogues

A mock photometric catalogue is then generated in order to mimic the VIDEO photometry, following the method presented in Laigle et al. (2019). We recall below the main aspects of the building of the catalogue. Galaxy fluxes are calculated from the total distribution of stellar particles in each galaxy using the stellar population synthesis (SPS) models from Bruzual & Charlot (2003) with a Chabrier (Chabrier 2003) initial mass function (IMF). Each star particle is assumed to behave as a single stellar population, and the total galaxy flux is the sum of the contributions of all star particles. Dust attenuation is implemented along the line of sight of each star particle, using the  $R_V = 3.1$  Milky Way dust grain model by Weingartner & Draine (2001). The gas metallicity distribution around the galaxies is taken as a proxy for dust distribution.

Galaxy spectra are then shifted according to the galaxy redshift, and convolved with the same filter passbands as the real data (namely  $u^*$ ,  $g'$ ,  $r'$ ,  $i'$ , and  $z'$  from CFHTLS, and  $Z$ ,  $Y$ ,  $J$ ,  $H$ , and  $K_s$  from VIDEO). Photometric errors are added in each band to reproduce the  $S/N$  distribution and sensitivity limit of VIDEO and CFHTLS filters (see Section 2). From this mock photometry, photometric redshifts and stellar masses are computed using the code LEPHARE

<sup>2</sup>Although we do not use it in this work, a twin simulation was also run without AGN feedback, the Horizon-no-AGN simulation, making it particularly useful for consistently testing the role of AGNs in galaxy formation (Peirani et al. 2017; Beckmann et al. 2017).



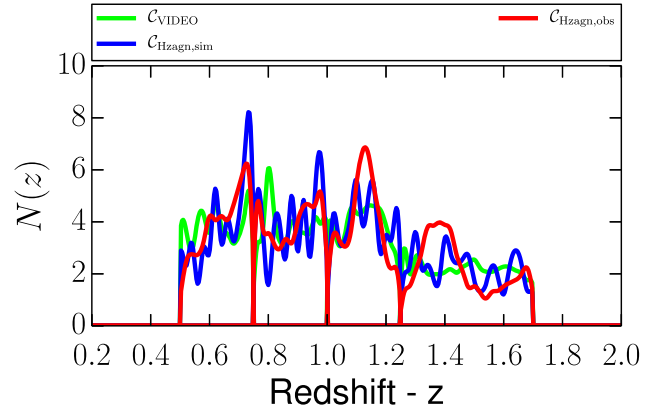


**Figure 1.** Distribution of the stellar masses and redshifts (blue points) of all  $K < 23.5$  galaxies in  $C_{\text{Hzagn,obs}}$ . The red lines denote the bins used in Hatfield et al. (2016). The red points mark the stellar mass limit for all objects that could be detected with the apparent magnitude limit of  $K_s < 23.5$ , and the green curve the implied 90 per cent stellar mass completeness limit, following the approach of Johnston et al. (2015). The green line matches the bottom edge of the red bins, as a confirmation that the mock catalogue has similar mass limits as  $C_{\text{VIDEO}}$ .

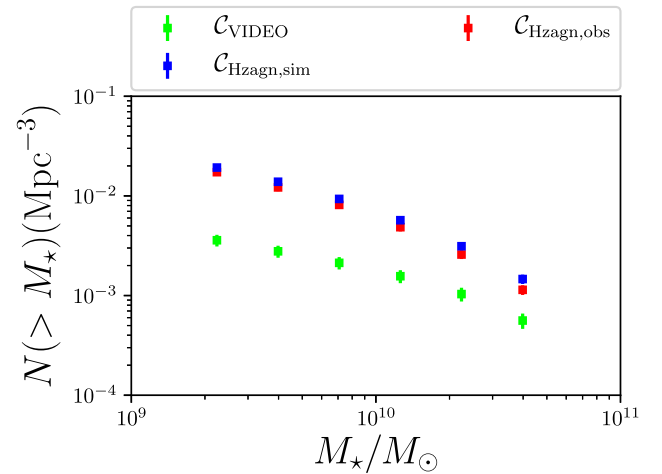
(Arnouts et al. 2002; Ilbert et al. 2006) with a configuration similar to Laigle et al. (2019).

Two simulated catalogues are used in this work. One, called  $C_{\text{Hzagn,sim}}$  in the following, contains the intrinsic stellar masses and redshifts. The other,  $C_{\text{Hzagn,obs}}$ , uses stellar masses and redshifts derived from the mock photometry through SED fitting and therefore naturally incorporates the systematics arising when computing physical properties from their photometry. A detailed study of these systematics has been presented in Laigle et al. (2019). In the following (Section 6.1), we will in particular investigate how the systematics propagate in the clustering measurements. Note that we incorporate observational limitations only up to a point. In particular, the galaxy photometry is derived from the entire distribution of particles, and not from a two-dimensional extraction on realistic images. Although noise is added to the mock photometry in a statistical way, it does not include systematics that usually degrade the observed photometry, due to blending, object fragmentation, variable point spread function (PSF), etc. Furthermore, as discussed in Laigle et al. (2019), the mock photometry includes less variety than the real one, because of the use of a single and constant IMF, a single SPS model, a single dust attenuation law, and the absence of nebular emission lines. As a consequence, the uncertainties on the photometric redshifts and stellar masses might be underestimated, as might be the amplitude of the systematics in the clustering measurements we derive in the following sections.

Fig. 1 (cf. fig. 1 in Hatfield et al. 2016) shows the  $C_{\text{Hzagn,obs}}$  masses and redshifts, after having removed galaxies fainter than  $K_s > 23.5$  (based on their mock photometry) as we did with the observations, leaving 259 567 galaxies (more than in VIDEO because of overproduction of low-mass galaxies in Horizon-AGN; see Section 6.2). We apply the same criteria as Hatfield et al. (2016) from Pozzetti et al. (2010) and Johnston et al. (2015) to determine the 90 per cent completeness limit. This limit is found by calculating the lowest stellar mass that each galaxy could have been detected at, using  $\log_{10}(M_{\text{lim}}) = \log_{10}(M_*) + 0.4(K_s - K_{\text{lim}})$  (where  $M_{\text{lim}}$  is this lowest mass that the galaxy could have been observed at,  $M_*$  is the galaxy mass,  $K_s$  is the galaxy  $K_s$ -band magnitude, and  $K_{\text{lim}}$  is the  $K_s$ -band limit used,  $K_{\text{lim}} = 23.5$  here). We can then find the 90th percentile of these stellar mass limit values, as a function of redshift, which corresponds to a 90 per cent completeness limit. This limit



**Figure 2.** The redshift distributions of  $C_{\text{Hzagn,obs}}$ ,  $C_{\text{Hzagn,sim}}$ , and  $C_{\text{VIDEO}}$ , each catalogue and each redshift bin normalized to have an area of 1.



**Figure 3.** The number counts from  $C_{\text{Hzagn,obs}}$ ,  $C_{\text{Hzagn,sim}}$ , and  $C_{\text{VIDEO}}$  for the  $0.5 < z < 0.75$  redshift bin.

is very similar to that found for the observations (as expected), allowing us to consistently compare all the VIDEO sub-samples to the Horizon-AGN mock catalogue.

We also confirm that the redshift distributions within the redshift bins are relatively similar between the two Horizon-AGN catalogues and the VIDEO catalogue (Fig. 2). The two Horizon-AGN catalogues have very similar redshift distributions (as expected); the medians of the redshift distributions are within 0.002 of each other for the first three redshift bins, and 0.014 different in the fourth redshift bin (the bin with the largest range). Similarly, the medians of the redshift distributions in each VIDEO redshift bin are within 0.02 of the Horizon-AGN in the first three redshift bins, and 0.04 in the fourth redshift bin. In summary, in terms of comparing like to like, in all our analysis we are comparing comparable redshift distributions, with the possible very minor exception of the  $1.25 < z < 1.7$  VIDEO bin, which has a slightly higher median redshift than the corresponding Horizon-AGN samples. Finally, in Fig. 3 we also show a sample comparison of the number counts in the three catalogues considered in this work ( $C_{\text{Hzagn,obs}}$ ,  $C_{\text{Hzagn,sim}}$ , and  $C_{\text{VIDEO}}$ ) as a function of stellar mass. See also fig. A3 in Laigle et al. (2019). The  $C_{\text{Hzagn,obs}}$  and  $C_{\text{Hzagn,sim}}$  one-point statistics are very similar (note the difference is better interpreted as a shift in the  $x$ -axis from a bias in the stellar mass estimates, rather than a

y-axis shift in the actual number of galaxies). The  $C_{\text{VIDEO}}$  number counts are substantially different from those in the Horizon-AGN catalogues due to different physics in the simulation compared to that in the real Universe – discussed in greater depth in Section 6.2

## 4 METHODS

In this work, we apply the same methodology as used for  $C_{\text{VIDEO}}$  (Hatfield et al. 2016) for calculating the angular correlation function and fitting HOD models in both simulated catalogues  $C_{\text{Hzagn,sim}}$  and  $C_{\text{Hzagn,obs}}$ .

Note that unlike in real data sets, there is no need to remove stars and no mask is needed for the simulated data. For  $C_{\text{Hzagn,obs}}$ , the redshifts have uncertainties estimated from LEPHARE and we use the weighting system of Arnouts et al. (2002) when calculating the correlation function; for  $C_{\text{Hzagn,sim}}$ , we simply use the true redshifts for each source (equivalently we give each source a weight of 1 for the redshift range it is in, and 0 for the other redshift bins). The VIDEO field and the Horizon-AGN light-cone are both  $1 \text{ deg}^2$  fields, so the integral constraint (a bias to measurements of the correlation function due to finite field effects) affects all measurements in a near-identical manner.

### 4.1 The two-point correlation function

The angular two-point correlation function  $\omega(\theta)$  is a measure of how much more likely it is to find two galaxies at a given angular separation than in a uniform Poissonian distribution:

$$dP = \sigma(1 + \omega(\theta))d\Omega, \quad (1)$$

where  $dP$  is the probability of finding two galaxies at an angular separation  $\theta$ ,  $\sigma$  is the surface number density of galaxies, and  $d\Omega$  is solid angle. We require  $\omega(\theta) > -1$  and  $\lim_{\theta \rightarrow \infty} \omega(\theta) = 0$  for non-negative probabilities and for non-infinite surface densities, respectively. The two-point correlation function is constructed to not be a function of number density (although you might expect them to be related for physical reasons).

The conventional way to estimate  $\omega(\theta)$  for the autocorrelation function is with the Landy & Szalay (1993) estimator, which is based on calculating  $\text{DD}(\theta)$ , the normalized number of galaxies at a given separation in the real data,  $\text{RR}(\theta)$ , the corresponding figure for a synthetic catalogue of random galaxies identical to the data catalogue in every way (i.e. occupying the same field) except position, and  $\text{DR}(\theta)$ , the number of galaxy to synthetic point pairs:

$$\omega(\theta) = \frac{\text{DD} - 2\text{DR} + \text{RR}}{\text{RR}}. \quad (2)$$

### 4.2 Halo occupation distribution modelling

Halo occupation modelling is a popular way of modelling galaxy clustering measurements, and has been shown to give physical results in agreement with other methods (e.g. Coupon et al. 2015; Chiu et al. 2016; see Coupon et al. 2012 and McCracken et al. 2015, for a more complete breakdown). A given set of galaxy occupation statistics is given, usually parametrized by three to five numbers, e.g. the number of galaxies in a halo as a function of halo mass. The model correlation function is broken down to a ‘one-halo’ term, describing the small-scale clustering of galaxies within an individual halo, and a ‘two-halo’ term, describing the clustering of the haloes themselves. The ‘one-halo’ term is constructed by convolving the profile of galaxies within a halo with itself, weighting

by the number of galaxies in the halo, and then integrating over all halo masses. The profile is usually taken to be one galaxy at the centre of the halo (the ‘central’) and all other galaxies tracing a Navarro–Frank–White (NFW; Navarro, Frenk & White 1996) profile. The two-halo term is constructed by scaling the DM linear correlation function by the weighted-average halo bias of the host haloes. The transition from the non-linear one-halo term to the linear two-halo term occurs at  $\sim 0.02\text{--}0.05^\circ$  (i.e. within the angular scales we measure clustering for) over the redshifts considered in this work.

The most general HOD parametrization commonly used is that of Zheng et al. (2005), which gives the total number of galaxies in a halo as

$$\langle N_{\text{tot}}(M_h) \rangle = \langle N_{\text{cen}}(M_h) \rangle + \langle N_{\text{sat}}(M_h) \rangle, \quad (3)$$

the total number of central galaxies as

$$\langle N_{\text{cen}}(M_h) \rangle = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{\log_{10} M_h - \log_{10} M_{\text{min}}}{\sigma_{\log M}} \right) \right), \quad (4)$$

and the total number of satellites as

$$\langle N_{\text{sat}}(M_h) \rangle = \langle N_{\text{cen}}(M_h) \rangle \left( \frac{M_h - M_0}{M_1} \right)^\alpha \quad (5)$$

(when  $M_h > M_0$ ; otherwise, the number of satellites is zero). This model has five parameters:  $M_{\text{min}}$  describes the minimum halo mass required to host a central galaxy,  $\sigma_{\log M}$  describes how sharp this step jump is (equivalent to the central to halo mass scatter),  $M_0$  is a halo mass below which no satellites are found, and  $M_1$  is the scale mass at which the halo begins to accumulate satellites. The power-law index  $\alpha$  describes how the number of satellites grows with halo mass.

### 4.3 MCMC fitting

To model our angular correlation functions, we compare to model correlation functions from the HALOMOD<sup>3</sup> code (Murray, Power & Robotham, in preparation). First, a spatial correlation function is calculated, which is then projected to an angular correlation function (as per Limber 1954), using a redshift distribution derived by smoothing the true redshift values of the galaxies for the  $C_{\text{Hzagn,sim}}$  correlation function, and the weighting system of Arnouts et al. (2002) for the  $C_{\text{Hzagn,obs}}$  correlation function. We then subtract off the numerical approximation of the integral constraint to get our final model correlation function.

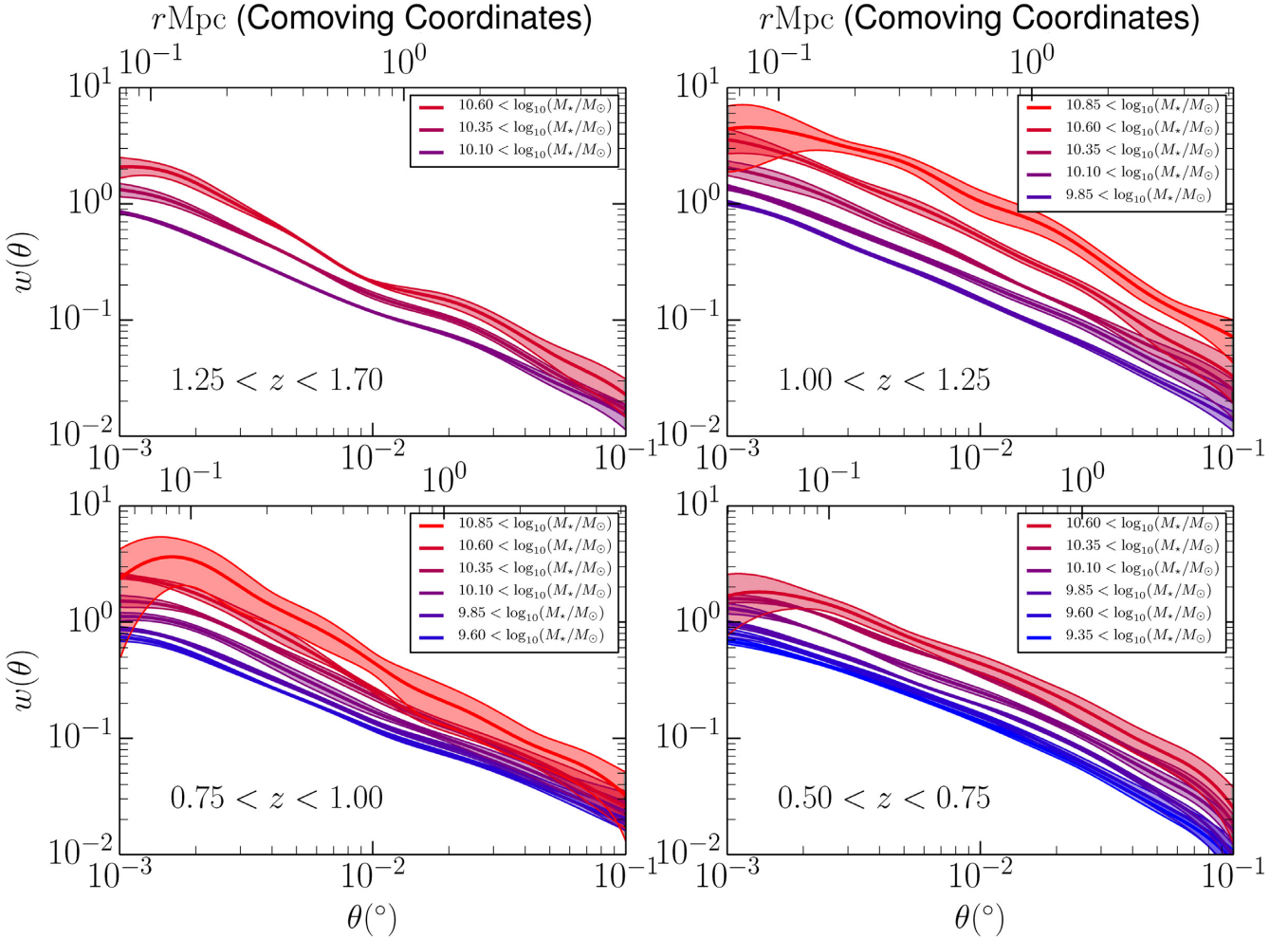
We use EMCEE<sup>4</sup> (Foreman-Mackey et al. 2013) to provide a Markov chain Monte Carlo (MCMC) sampling of the parameter space to fit our correlation function. The likelihood is defined by

$$\chi^2 = \frac{[n_{\text{gal}}^{\text{obs}} - n_{\text{gal}}^{\text{model}}]^2}{\sigma_n^2} + \sum_i \frac{[\omega^{\text{obs}}(\theta_i) - \omega^{\text{model}}(\theta_i)]^2}{\sigma_{\omega_i}^2}, \quad (6)$$

where  $n_{\text{gal}}^{\text{obs}}$  is the  $C_{\text{Hzagn,sim}}/C_{\text{Hzagn,obs}}$  galaxy number density,  $n_{\text{gal}}^{\text{model}}$  is the HOD model galaxy number density,  $\sigma_n$  is the error on the number density including both Poisson noise and cosmic variance (calculated as per Trenti & Stiavelli 2008),  $\theta_i$  are the angular scales we fit over,  $\omega^{\text{obs}}$  is the observed angular correlation function,  $\omega^{\text{model}}$  is the angular correlation function of a given HOD model, and  $\sigma_{\omega_i}$  is the error on the measurements of the correlation function from bootstrapping of the data.

<sup>3</sup><https://github.com/steven-murray/halomod>

<sup>4</sup><http://dan.iel.fm/emcee/current/>



**Figure 4.** The angular correlation function for each stellar mass and redshift bin, measured in the simulated catalogue  $C_{\text{Hzagn},\text{sim}}$  in different redshift bins. The dotted lines represent  $1\sigma$  uncertainties on the measurements. The comoving scale on the upper  $x$ -axes corresponds to the median redshift of the samples.

We use a uniform prior over  $10 < \log_{10}(M_{\text{min}}/M_{\odot}) < 15$ ,  $\log_{10}(M_{\text{min}}/M_{\odot}) < \log_{10}(M_1/M_{\odot}) < 17$ ,  $8 < \log_{10}(M_0/M_{\odot}) < \log_{10}(M_1/M_{\odot})$  (uniform in log space),  $0.5 < \alpha < 2.5$ , and  $0 < \sigma < 0.6$ . We used 20 walkers with 1000 steps, which have starting positions drawn uniformly from the prior.

We use 500 000 random data points in this study. We use 100 bootstrap resamplings to estimate the uncertainty at the 16th and 84th percentiles of the resampling. For  $n_{\text{gal}}^{\text{obs}}$ , for  $C_{\text{Hzagn},\text{sim}}$  we use the number of galaxies divided by the appropriate volume, and for  $C_{\text{Hzagn},\text{obs}}$  we use the sum of the weights as for  $C_{\text{VIDEO}}$  (Hatfield et al. 2016).

## 5 RESULTS

### 5.1 ACF measurements in Horizon-AGN

The clustering measurements from  $C_{\text{Hzagn},\text{sim}}$ , i.e. without any observational errors, are shown in Fig. 4. The behaviour usually measured in observational data sets is correctly recovered.

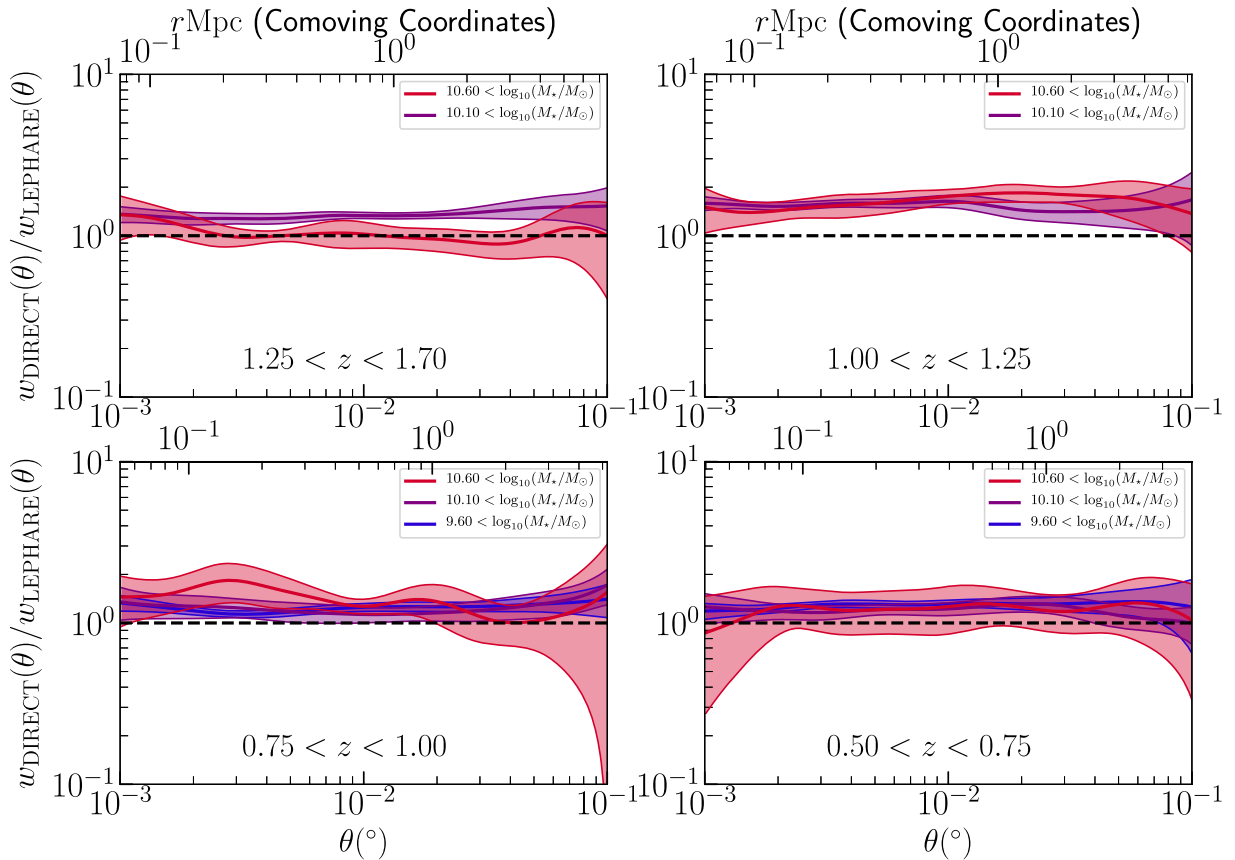
The mock correlation functions have near power-law behaviour, with sensible amplitudes increasing with stellar mass at all redshifts. To some degree, this is expected – the underlying DM distribution is robust between different simulations, and bias is

expected to be  $\lesssim 2-3$ , so it would be extremely surprising if it was orders of magnitude greater or less than expected. However, the agreement over linear and non-linear scales should reassure us that Horizon-AGN is correctly qualitatively capturing the large-scale distribution of galaxies. The clustering measurements from  $C_{\text{Hzagn},\text{obs}}$  look very similar; how they compare to  $C_{\text{Hzagn},\text{sim}}$  is discussed in Section 5.2. The exception is the highest stellar mass bin at  $0.75 < z < 1$  in the LEPHARE mock catalogue, for which there were too few galaxies to measure the clustering – this  $C_{\text{Hzagn},\text{obs}}$  bin is excluded in the subsequent analyses.

### 5.2 Impact of observational uncertainties on clustering measurement

In Fig. 5, we compare the clustering measurements from the mock catalogues  $C_{\text{Hzagn},\text{sim}}$  and  $C_{\text{Hzagn},\text{obs}}$  for the same stellar mass and redshift bins when the sources are binned by (a) their true redshift and stellar mass within the simulation and (b) their photometric redshift and stellar mass (as derived from the simulated fluxes, and using the same weighting scheme as was applied to  $C_{\text{VIDEO}}$ ), respectively.

The two sets of clustering measurements agree to moderate accuracy, never differing by more than a factor of 2, typically with



**Figure 5.** Ratio of the Horizon-AGN clustering measurements for the stellar mass and redshift bins when (a) taking the stellar masses and redshifts directly from the simulation ( $\mathcal{C}_{\text{Hzagn,sim}}$ ) and (b) taking the stellar masses and redshifts from the LEPHARE-based catalogues ( $\mathcal{C}_{\text{Hzagn,obs}}$ ). Only alternating stellar mass bins are plotted for clarity of plot. The thin lines are the errors on the ratio, with the errors from both sets of Horizon-AGN clustering measurements propagated through. The dashed line corresponds to  $\omega_{\text{DIRECT}} = \omega_{\text{LEPHARE}}$ .

comparatively little angular dependence. The  $\mathcal{C}_{\text{Hzagn,sim}}$  clustering measurement is also consistently higher than inferred from  $\mathcal{C}_{\text{Hzagn,obs}}$  because of scattering between stellar mass and redshift bins, which inevitably reduces the clustering signal. Although weak, there is slight trend for this systematic reduction to increase with increasing redshift, as redshift uncertainties tend to be higher at higher redshift.

### 5.3 Comparing VIDEO and Horizon-AGN

We compare the  $\mathcal{C}_{\text{Hzagn,obs}}$  results to the corresponding VIDEO subsample measurements from Hatfield et al. (2016). In Fig. 6, we compare the  $\mathcal{C}_{\text{Hzagn,obs}}$  clustering measurements, as opposed to the  $\mathcal{C}_{\text{Hzagn,sim}}$  clustering measurements in order to compare ‘like for like’ since the VIDEO results are also derived from photometry and have similar biases. In particular, we show the ratio of the VIDEO and Horizon-AGN angular correlation functions in Fig. 6. Note that the one-point statistics of these samples, at the same stellar masses, are quite different (Fig. 3), which should be borne in mind when interpreting these data.

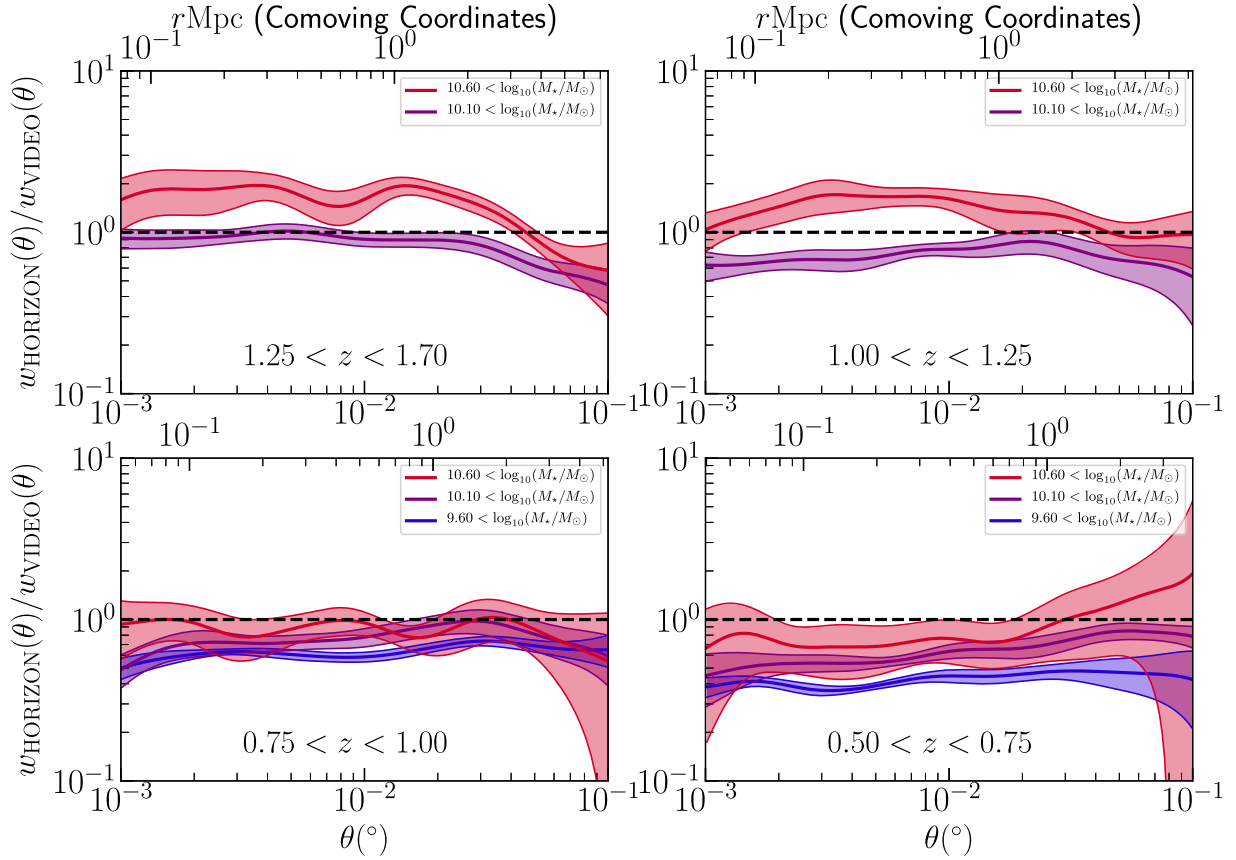
The Horizon-AGN clustering measurements are greater or lesser than the VIDEO measurements by at most a factor of 3 in any bin. In general, the trend in each redshift bin is that  $\omega_{\text{Horizon}}/\omega_{\text{VIDEO}}$  increases with stellar mass; e.g. at lower masses, Horizon-AGN increasingly underestimates the clustering, indicating that clustering in Horizon-AGN has too strong a dependence on stellar mass. At

lower redshifts, Horizon-AGN underestimates the clustering, and at higher redshifts is closer to the VIDEO clustering, sometimes starting to slightly overestimate it. The underestimation (particularly at low redshifts and stellar masses) has very high statistical significance ( $>5\sigma$ ); however, the overestimations (at higher stellar masses and redshifts) typically have lower ( $1-2\sigma$ ) significance. In most bins, there is very little angular scale dependence with the possible minor exception of large scales at high redshift – possibly a consequence of that bin in Horizon-AGN having low power at large scales.

### 5.4 HOD fits

Fig. 7 shows the HOD  $M_{\text{min}}$  and  $M_1$  parameters (characteristic halo masses required to host central and satellite galaxies, respectively). This is plotted as a function of stellar mass for the different redshift bins, as inferred from  $\mathcal{C}_{\text{Hzagn,sim}}$  and from  $\mathcal{C}_{\text{Hzagn,obs}}$  (both in Horizon-AGN). The parameters are largely very similar when inferred from the direct catalogue and the LEPHARE catalogue. We do not show the plots here, but the  $\alpha$  values for both catalogues agree within uncertainties, and are essentially consistent with  $\alpha = 1$  (as per the results in Hatfield et al. 2016 for the VIDEO observations). The  $M_0$  values also agree, but again similarly to Hatfield et al. (2016) are very poorly constrained (as it describes a cut-off in the number of satellites at low halo masses where the expected number of satellites is already much less than 1).





**Figure 6.** Ratio of the VIDEO and  $C_{\text{Hzagn,obs}}$  clustering measurements for the stellar mass and redshift bins. Only alternating stellar mass bins are plotted for clarity of plot. The shaded area covers the uncertainties on the ratio, with the uncertainties on both the VIDEO and Horizon-AGN measurements propagated through. The dashed line corresponds to  $\omega_{\text{Horizon}} = \omega_{\text{VIDEO}}$ .

Fig. 8 shows the HOD  $\sigma$  parameter as a function of stellar mass for the different redshift bins, as inferred from  $C_{\text{Hzagn,sim}}$  compared to those from  $C_{\text{Hzagn,obs}}$  (both in Horizon-AGN). The scatter inferred when observational uncertainties are included is significantly higher:  $\sim 0.6$  typically measured for  $C_{\text{Hzagn,obs}}$  and  $\sim 0.3\text{--}0.5$  typically measured for  $C_{\text{Hzagn,sim}}$ . Note the use of a uniform prior over  $[0, 0.6]$  for  $\sigma$  as per McCracken et al. (2015) and Hatfield et al. (2016) is now substantiated by the knowledge that observational systematics bias scatter estimates to higher values. The true scatter in the simulation is even lower,  $\sim 0.2$ , so even without the observational systematics we still overestimate the scatter by about 50 per cent through the fitting process.

We show in Fig. 9 the best-fitting  $M_{\text{min}}$  and  $M_1$  values from the HOD models fit to the LEPHARE Horizon-AGN measurements, alongside the equivalent VIDEO measurements from Hatfield et al. (2016). The Horizon-AGN fits give the same qualitative behaviour as observed in VIDEO,  $M_{\text{min}}$  and  $M_1$  growing as approximate power laws with stellar mass, with  $M_1$  roughly an order of magnitude larger than  $M_{\text{min}}$ . Horizon-AGN measurements for both  $M_{\text{min}}$  and  $M_1$  are generally slightly lower than the equivalent VIDEO measurements.  $M_{\text{min}}$  values agree well at high stellar masses ( $\sim 10^{10.85} M_{\odot}$ ), but Horizon-AGN  $M_{\text{min}}$  values are around 0.5 dex lower than the VIDEO values for stellar masses  $\sim 10^{9.6} M_{\odot}$ . This is consistent with Fig. 6, where the clustering amplitude in Horizon-AGN is much lower than that measured in VIDEO at lower stellar masses.  $M_1$  values are consistently a factor of  $\sim 2$  lower for Horizon-AGN

than for VIDEO. Similarly to in our VIDEO measurements, there appears to be very little evolution in  $M_{\text{min}}$  and  $M_1$  as a function of redshift, over the redshift and stellar mass ranges considered here.

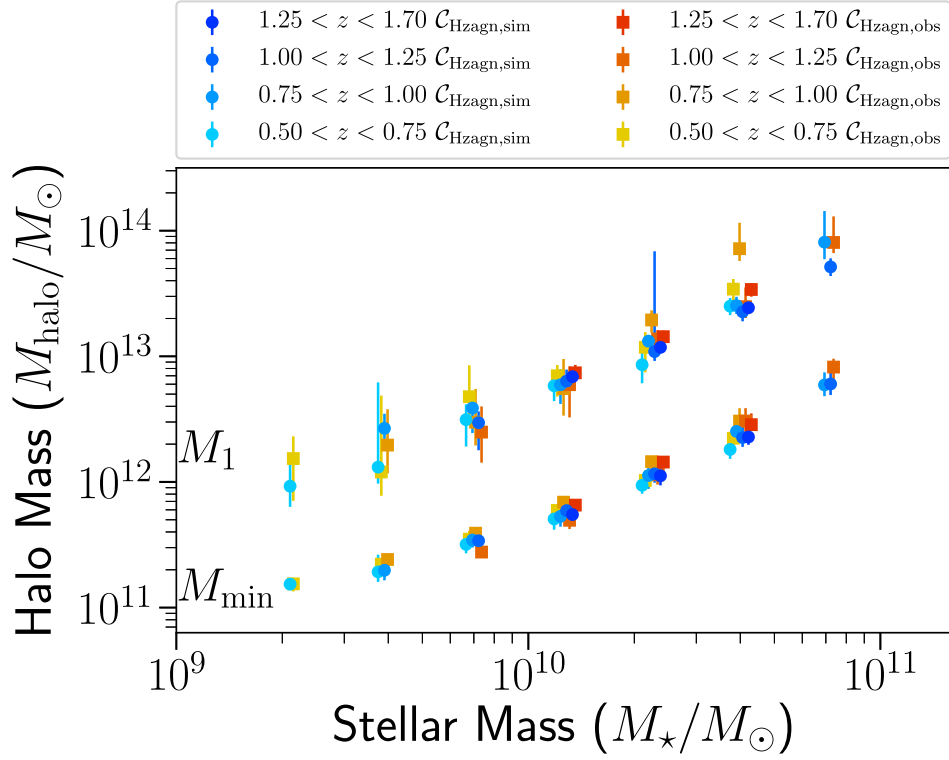
## 6 DISCUSSION

Causes of discrepancies between observed and simulated measurements can be divided into three categories. First, discrepancies can arise from the simulation inherently not completely capturing the correct galaxy formation physics; e.g. the simulation genuinely creates more galaxies of a given mass than are present in the real Universe. Secondly, discrepancies can arise when observational biases are poorly understood and accounted for; e.g. the incompleteness of a galaxy survey is underestimated, leading us to falsely conclude that the simulation overproduces galaxies, even if the simulation actually matches reality. Finally, there can be differences greater than expected from statistical uncertainty from *cosmic variance* – the galaxy formation physics could be correctly captured in the simulation, but large-scale structure variance could still make measurements from the simulation differ from observations non-trivially.

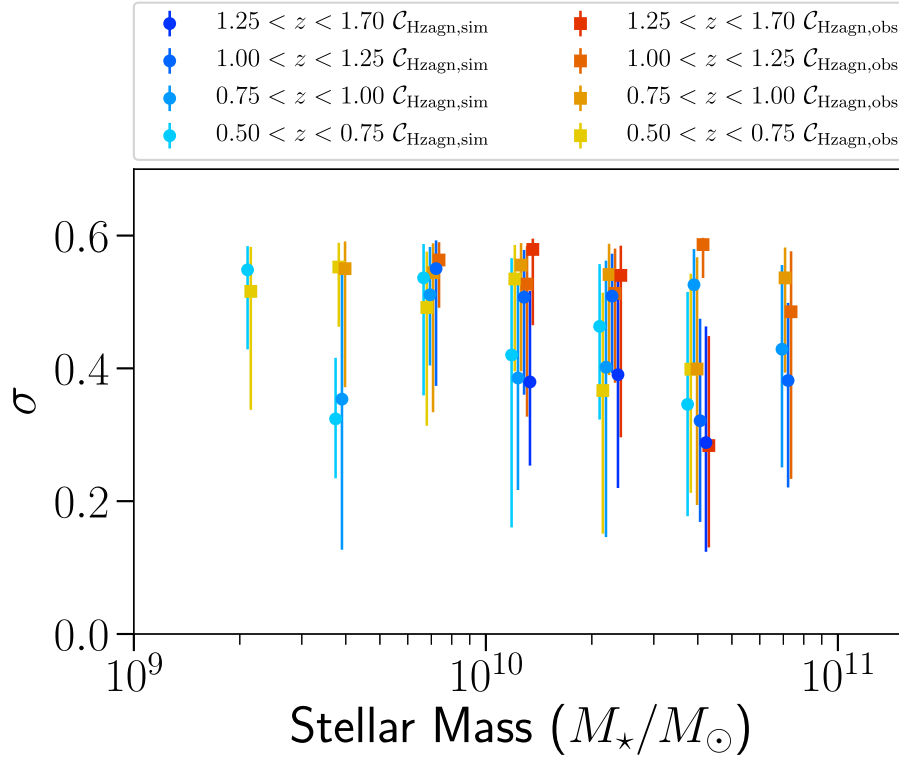
### 6.1 Observational biases

A key complexity in comparing results from observations and simulations is that the two ‘tools’ perceive galaxies in a dramatically

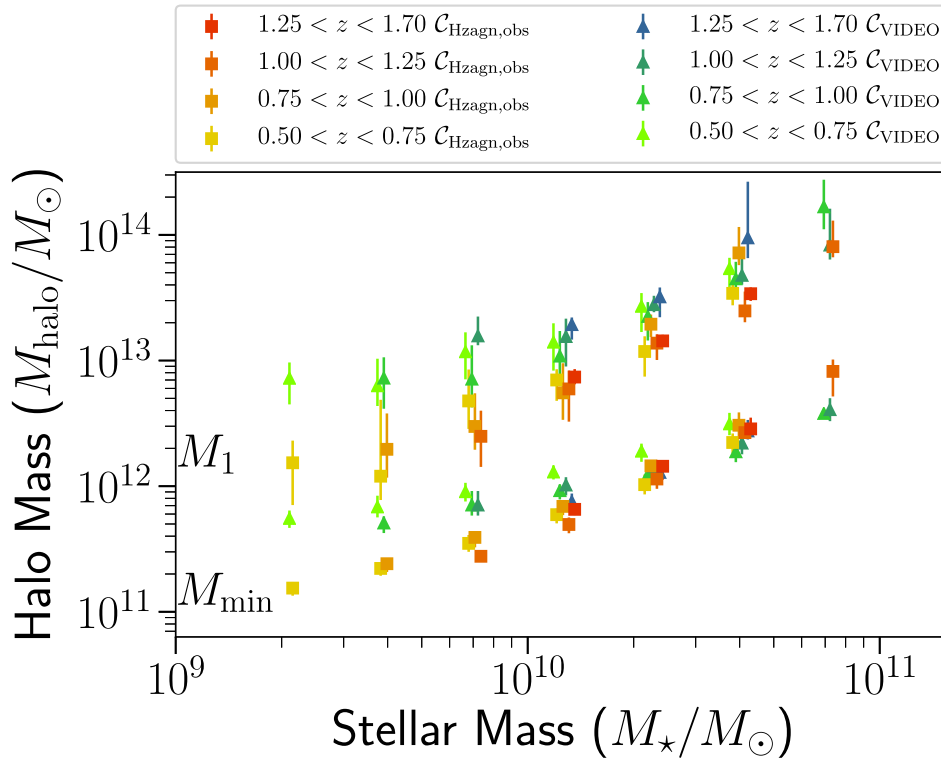




**Figure 7.** Best-fitting  $M_{\min}$  and  $M_1$  for the Horizon-AGN clustering measurements, as a function of galaxy sample threshold stellar mass, for both  $\mathcal{C}_{\text{Hzagn,sim}}$  and  $\mathcal{C}_{\text{Hzagn,obs}}$ . Data are plotted with small stellar mass offsets for clarity of plot.



**Figure 8.** Best-fitting  $\sigma$  values for the Horizon-AGN clustering measurements, as a function of galaxy sample threshold stellar mass, for both  $\mathcal{C}_{\text{Hzagn,sim}}$  and  $\mathcal{C}_{\text{Hzagn,obs}}$ . Data are plotted with small stellar mass offsets for clarity of plot.



**Figure 9.** Best-fitting  $M_{\min}$  and  $M_1$  for the Horizon-AGN  $C_{\text{Hzagn,obs}}$  clustering measurements, as a function of galaxy sample threshold stellar mass. Equivalent measurements in VIDEO from Hatfield et al. (2016) shown for comparison. Data are plotted with small stellar mass offsets for clarity of plot.

different manner. Simulations are concerned with (stellar) mass, whereas observations deal with luminosity. Comparisons between observations and simulations inherently require either inferring mass from observed luminosities or predicting a luminosity from a simulated mass.

Laigle et al. (2019) fit a library of templates to the Horizon-AGN mock photometry (essentially treating the photometry as if it was observations) to investigate whether the true galaxy masses and redshifts within the simulation can be recovered correctly from the photometry. In the COSMOS configuration, they find a scatter of  $\sim 0.1$  dex between the intrinsic and photometric stellar masses (see their fig. 6). More problematically, they also find (for the redshifts relevant to this work) that the inferred stellar masses are biased to values at most  $\sim 0.12$  dex less than the true values. This estimate holds for the COSMOS configuration, i.e. for redshifts and masses derived from the fitting of 30 photometric bands. However, even if the photometric baseline used in our analysis contains less photometric bands, the NIR photometric range (which matters the most for stellar mass computation) is still very well sampled with the VIDEO observations. Therefore, the performance of the VIDEO configuration in terms of the precision of the reconstructed stellar mass is very similar to the COSMOS one. Clearly, the systematic underestimation of the mass is relevant for consistently comparing clustering results between observations and simulations. If it is the case that observationally inferred stellar masses are consistently 0.1 dex less than the true values, then comparing the clustering of  $M_* > 10^{10} M_\odot$  observed galaxies with clustering of  $M_* > 10^{10} M_\odot$  simulated galaxies is really comparing the clustering of  $M_* > 10^{10.1} M_\odot$  observed galaxies to  $M_* > 10^{10} M_\odot$  simulated galaxies. The higher mass sample will typically have a higher clustering amplitude, so underestimating the stellar masses

of observed galaxies will falsely make it appear that the simulation underestimates clustering amplitude (even if it correctly captures galaxy physics). Scatter in the stellar mass conversely will reduce clustering amplitude, as the more numerous lower stellar mass galaxies (in lower mass haloes) will be preferentially ‘up-scattered’ into higher stellar mass bins. Photometric redshift uncertainties will ordinarily have comparatively little impact when the scatter is within an individual redshift bin, but scatter between redshift bins will typically reduce the clustering amplitude, as structure at redshifts with large separations is largely uncorrelated.

Comparison of the HOD fits from the  $C_{\text{Hzagn,sim}}$  and  $C_{\text{Hzagn,obs}}$  catalogues allows us to investigate directly the impact of using photometric stellar masses and redshifts instead of the true value, and develops the discrepancy highlighted in Fig. 5. In particular, the scatter in stellar mass itself induces a higher measured  $\sigma$  [effectively scatter in central stellar mass to halo mass ratio (SMHR)] – the  $C_{\text{Hzagn,obs}}$  scatter is about 0.1–0.2 dex greater than the  $C_{\text{Hzagn,sim}}$  scatter (see Fig. 8), which is comparable to the scatter in fig. 6 in Laigle et al. (2019). In other words, the measured scatter in SMHR should be interpreted as the sum of both the intrinsic and the photometric scatter. The  $\sigma$  measured in VIDEO observations in Hatfield et al. (2016) is also  $\sim 0.6$  – our results would suggest that this is likely an overestimate resulting from scatter in the estimates of the stellar mass.

Similarly, the marginally higher  $M_{\min}$  values for  $C_{\text{Hzagn,obs}}$  correspond to the bias in the estimate of the stellar mass; the  $C_{\text{Hzagn,obs}}$  points should essentially be slightly shifted to the right in Figs 7 and 9. In summary, the reduced clustering amplitude from photometric systematics described in Section 5.2 mainly increases our measurements of the SMHR scatter in HOD modelling, as opposed as biasing the estimated halo masses themselves.

### 6.1.1 Limitations of our modelling

Our estimates of the systematics in the amplitude of the clustering must be taken as an optimistic lower limit. As briefly described in Section 3 and in more detail in Laigle et al. (2019), there are two main limitations of our modelling of observational uncertainties.

First, the mock photometry of the simulated galaxies includes less variety than the real photometry (mainly due to the use of constant IMF and SPS models, a constant dust-to-metal ratio, and single dust attenuation law, and no modelling of the emission lines). For example, fitting the photometry with an IMF different from that used to produce this photometry might cause additional systematics of the order of 0.1 dex.<sup>5</sup>

Secondly, although photometric errors are included in a statistical sense in our mock photometry, we do not model any of the possible systematics occurring at the stage of photometric extraction. These systematics are numerous and might include, in particular, PSF variations over the field and as a function of wavelengths, blending of nearby galaxies or conversely fragmentation of galaxies with perturbed morphologies, errors in the astrometry, and imperfect removal of the background. The only way to quantify consistently the importance of all these additional sources of noise is to perform an end-to-end extraction of the photometry directly from mock images, tuned to incorporate the characteristics of the instrument, which will be presented in a future work. Most of these effects are, however, likely to be important especially at faint magnitudes, and our cut at  $K_s < 23.5$  should preserve the vast majority of galaxies in our sample from strong systematics. However, it must be noted that bright objects might still suffer from some of these effects, e.g. PSF variations or miscentring.

As a consequence of these two limitations in our pipeline, the redshift and masses uncertainties in  $C_{\text{Hagn,obs}}$  are likely to be underestimated, as outlined by fig. 4 in Laigle et al. (2019) (a comparison of COSMOS2015 and Horizon-AGN errors as derived from LEPHARE). We would anticipate that increasing the redshift uncertainties would further increase the underestimation of the clustering amplitude, and therefore the analysis presented here should be considered as an optimistic case.

## 6.2 Overproduction of lower mass galaxies in Horizon-AGN

Although observational biases may play some role,<sup>6</sup> the main discrepancy in the galaxy–halo connection from clustering between observations and simulations apparent in our results is that the SMHR does not drop off as fast at low masses in Horizon-AGN as in VIDEO observations (Fig. 9).

This raised SMHR for low-mass haloes in Horizon-AGN (equivalent to the overproduction of lower mass galaxies, or galaxies in lower mass haloes growing too large relative to observations) is reported in Dubois et al. (2014) and Kaviraj et al. (2017). Kaviraj et al. (2017) compare stellar mass functions from the literature to those directly from the simulation, finding an overabundance of galaxies with stellar masses  $\lesssim 10^{11} M_\odot$  (cf. Fig. 3). Dubois et al. (2014) compare the SMHR for central galaxies and their haloes directly from the simulation to observational results from abundance matching in Moster, Naab & White (2013), finding

that SMHR observations and simulations agree for haloes with masses  $\approx 10^{12} M_\odot$ , but that galaxies in  $\approx 10^{11} M_\odot$  haloes are approximately a factor of 10 more massive in Horizon-AGN than observations would suggest, which agrees with our results. Kaviraj et al. (2017) suggest that this discrepancy is due to missing physics in the sub-grid supernovae feedback prescription of the simulation. They propose that stronger regulation of star formation in lower mass haloes could be achieved either through more realistic treatment of the interstellar medium (Kimm et al. 2015) or through stronger winds from star clusters (Agertz & Kravtsov 2015). Note that the low halo mass regime is one in which using clustering as opposed to abundance matching is particularly important for inferring galaxy to halo mass ratios (see Sawala et al. 2015).

With simulations, it is possible to compare inferences made from mock observations to the true physics within the simulation; e.g. it is possible to compare the inferred halo masses from clustering to the true halo masses within the simulation. Fig. 10 shows the HOD inferences compared with the real halo masses in the simulation, and illustrates that the inferences from the clustering and HOD modelling essentially agree with the true galaxy–halo relation in the simulation. This suggests that the modelling in Hatfield et al. (2016) (and other similar works) is largely giving accurate inferences. The bias introduced to the estimates of halo masses from observational systematics is  $\sim 0.1$  dex (the difference between the inferences based on  $C_{\text{Hagn,sim}}$  and  $C_{\text{Hagn,obs}}$ ), and the bias introduced by the fitting process is  $\sim 0.05$ – $0.1$  dex<sup>7</sup> (the difference between the inferences from  $C_{\text{Hagn,sim}}$  correlation functions and the true physics within the simulation). The physics of the Horizon-AGN universe is, as discussed, different from that in the real Universe, so the sizes of systematics are not guaranteed to be exactly the same size for observational data, but this seems a plausible estimate of the sizes of systematics on real observations. This is particularly comforting given the comparative simplicity of HOD models; for example, (i) the model discussed here does not include assembly bias (see Zentner et al. 2014) and (ii) there is some evidence at low redshift that galaxies are not spherically symmetrically distributed in haloes (Pawlowski, Ibata & Bullock 2017). Pujol et al. (2017) in contrast found that SAMs and HOD models give different clustering predictions, mainly because of contributions of orphan galaxies and because the SAMs had different satellite galaxy radial profile densities than the NFW that are assumed for the HOD model (modifying the profiles is considered in Hatfield & Jarvis 2017, however). Hydrodynamical simulations do not have orphan galaxies per se, but they do capture some of the physics in a more fundamental manner; HOD models certainly will not capture everything within the simulation, but our results seem to show that they correctly capture the main features.

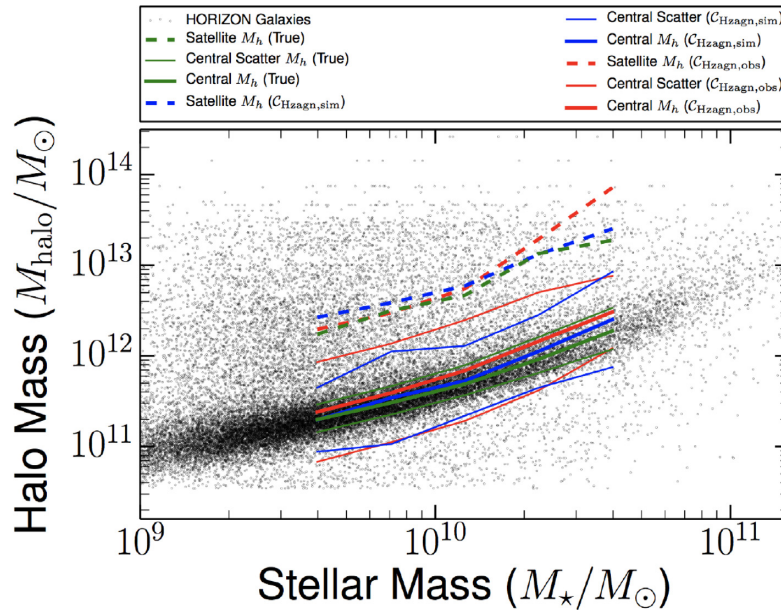
## 6.3 Cosmic variance

Cosmic variance refers to the observation that statistical measurements can vary by more than Poisson variance due to large-scale structure; e.g. the field might sample an extreme over- or underdensity of the Universe, which can impact on the statistical properties of the field. As discussed in Hatfield et al. (2016), there is already known to be moderate difference in the clustering properties of the VIDEO/CFHTLS-D1 field and the COSMOS field (Durkalec et al. 2015).

<sup>7</sup>With the exception of  $M_1$  for the highest stellar masses, where it gets up to  $\sim 0.5$  dex.

<sup>5</sup>Magnitudes are  $\sim 0.4$  mag fainter with a Salpeter IMF compared to a Chabrier one, so a mismatch between both IMFs at the SED fitting stage would cause a 0.1 offset in stellar mass.

<sup>6</sup>The Horizon-AGN LEPHARE catalogue and VIDEO have similar observational biases, but only to a certain degree, as discussed in Section 6.1.



**Figure 10.** Comparison of the stellar mass to host halo mass for galaxies taken directly from Horizon-AGN (black points), and stellar mass when central (filled line) with scatter (thin lines) and when satellites (dashed line), as inferred from clustering. The red curves show the inference from  $C_{\text{Hzagn,obs}}$ , the blue curves that from  $C_{\text{Hzagn,sim}}$ , and the green curves the true values within the simulation. Note that the scatter shown here is the inferred scatter on the central stellar mass to halo mass relation, not the uncertainty on this relation.

Cosmic variance-like effects, as described in Blaizot et al. (2005), can also impact mock observations derived from simulations in two main ways. First, only a finite volume of the Universe is simulated. This requires sampling initial conditions from some larger distribution – a different sampled set of initial conditions would lead to a different simulated universe. Secondly, the mock cone is constructed; for a given simulation, constructing the mock sky ‘from different viewpoints’ will give different sets of mock observations.

VIDEO has three separate fields, so in future work it should be possible to directly estimate the effects of cosmic variance on our observational results. To understand cosmic variance in the simulated results, ideally one would run the simulation multiple times, but this is currently unfeasible for hydrodynamical simulations<sup>8</sup> (although it is more viable in SAMs; e.g. Stringer et al. 2009). The approach of sampling different viewpoints within the same simulation (the second type of cosmic variance in simulations discussed earlier) may be more viable for hydrodynamic simulations, although as Blaizot et al. (2005) discuss, this will generally underestimate the true cosmic variance, as each generated mock sky will still be sampling the same density field, just from different angles.

Although to completely quantify cosmic variance multiple observational fields and multiple simulations with different initial conditions are needed, we can estimate the approximate size of cosmic variance if a few assumptions are made. In particular, as discussed, the galaxy–halo connection appears to have relatively little evolution with redshift. If we assume that the galaxy–halo connection is constant in all our redshift bins, and in addition we find it acceptable to treat the different redshift bins as independent volumes of space, we can use the differences between different

redshift bins to estimate the approximate size of the cosmic variance (as an alternative to using different fields). Different redshift bins are not completely independent (i) because of cross-bin contamination in the catalogues with observational biases and (ii) because adjacent redshift bins are essentially adjacent comoving volumes, and (iii) these comoving volumes are all ultimately drawn from the same simulation. Nonetheless, making these assumptions allows us to obtain a lower bound on the size of the cosmic variance on the parameters. We assume that the (non-systematic) uncertainty on our estimates of the HOD parameters can be expressed as  $\sigma_{\text{Total}}^2 = \sigma_{\text{CV}}^2 + \sigma_{\text{Stat}}^2$ , where  $\sigma_{\text{Total}}$  is the total uncertainty on the estimate,  $\sigma_{\text{CV}}$  is the cosmic variance on the estimate, and  $\sigma_{\text{Stat}}$  is the statistical uncertainty on the estimates calculated in Section 5. We then fit (using MCMC) a second-order polynomial in log–log space of the  $C_{\text{Hzagn,sim}}^9 M_{\text{min}}$  estimates as a function of  $M_*$ . When calculating the likelihood, we use Gaussian uncertainties of  $\sigma_{\text{Total}}$ , where the  $\sigma_{\text{Stat}}$  values are as calculated and  $\sigma_{\text{CV}}$  is unknown. This gives four unknowns: the three parameters of the polynomial model, and  $\sigma_{\text{CV}}$  – and we find  $\sigma_{\text{CV}} \gtrsim 0.05$  dex. This is an imperfect model, as cosmic variance is correlated within and between redshift bins, but likely gives a lower bound of the approximate size of the variance, in the absence of multiple fields and multiple simulations to analyse. We can also compare this estimate of the size of cosmic variance on our HOD measurements to the cosmic variance on one-point statistics, which we calculated using the method of Trenti & Stiavelli 2008 for equation (6). This approach uses assumptions about the two-point statistics (e.g. galaxy bias) to make estimates of the size of the cosmic variance of the one-point statistics. This method estimates the size of the cosmic variance on the counts to be  $\sim 0.05$  dex (slightly variable depending on redshift range under

<sup>8</sup>Although Pontzen et al. (2016) present an interesting way of capturing much of cosmic variance in hydrodynamical simulations using only two runs.

<sup>9</sup>The final estimate of rough size of the impact of cosmic variance is not strongly affected by using a different catalogue or slightly different ways of fitting the data, etc.



consideration, etc.). The cosmic variance on the clustering would depend on the three-point statistics, so will not be the same as the cosmic variance on the counts, and in addition uncertainty on the clustering does not translate linearly to uncertainty on HOD parameters. None the less, the fact that the first method gave a result comparable to the analytic result for number counts would suggest that  $\sim 0.05$  dex is probably a reasonable approximation of the size of cosmic variance on halo mass estimates.

## 7 CONCLUSIONS

In this paper, we investigated the clustering of galaxies in mock catalogues from the hydrodynamic cosmological simulation Horizon-AGN, to compare to observations and to confirm that HOD methods give correct physical deductions. We selected mock sources in the simulation that permitted direct comparison to VIDEO sub-samples described in Hatfield et al. (2016), using redshifts and stellar masses taken directly from the simulation as well as derived from the mock photometry. We then measured the angular correlation function and fitted HOD models using identical procedures to that applied to the observations so that a consistent comparison could be made.

We found that the correlation function measurements from the simulations were in qualitative agreement with the correlation functions observed in VIDEO, producing the correct approximate power-law behaviour, with the correct stellar mass dependence, and the amplitude differing by at most a factor of 3. HOD modelling of the Horizon-AGN measurements also qualitatively recovers the broad nature of the connection between galaxies and haloes suggested by VIDEO, namely,  $M_{\min}$  and  $M_1$  growing as power laws (with a factor of approximately an order of magnitude between them).

The (central) SMHR differs between Horizon-AGN and VIDEO, reinforcing the known result of Horizon-AGN that at low halo masses galaxies are more massive than observations would suggest (equivalently the stellar mass function overestimates the number of lower mass galaxies). HOD modelling recovers the galaxy–halo connection as taken directly from the simulation, suggesting that conventional HOD methodology is giving physically correct results.

Our key results are as follows:

- (i) The halo masses inferred from HOD modelling of clustering in mock Horizon-AGN catalogues closely match halo masses taken from the simulation directly, justifying our confidence that inferences from HOD modelling in observations are correct.
- (ii) Use of photometric redshifts and stellar masses biases clustering measurements (in particular reducing the amplitude of the angular correlation function by up to a factor of 2). This in turn biases the measurement of HOD parameters, with a relatively small impact on the inferred halo masses (essentially shifting the SMHR by the bias on the inferred stellar mass), but causing significant overestimates of the scatter in the inferred central SMHR. Non-statistical uncertainties on halo mass estimates in this study were  $\gtrsim 0.05$  dex from cosmic variance (dependent on field geometry),  $\sim 0.1$  dex from observational systematics (dependent on photometry used and data reduction process), and  $\sim 0.05$ – $0.1$  dex from systematics in the fitting process (dependent on modelling used).
- (iii) Clustering measurements in Horizon-AGN and the VIDEO survey observations disagree in a way consistent with the known overproduction of low-mass ( $M_* \lesssim 10^{10.5} M_{\odot}$ ) galaxies in Horizon-AGN, but agree for other stellar masses.

The analysis presented in this paper is based on  $1 \text{ deg}^2$ , and as we discussed, cosmic variance could play an important role in

driving the discrepancy between the simulation and VIDEO. In order to test how much VIDEO is sensitive to cosmic variance, future work will present an analysis based on the full  $12 \text{ deg}^2$  of the survey. In terms of our understanding of galaxy physics, future work will also investigate the potential need for stronger regulation of star formation in low-mass haloes in Horizon-AGN, or what other modifications to the simulation could help reduce the discrepancy between clustering observations and predictions. At the high-mass end, as suggested in Hatfield & Jarvis (2017), AGN feedback likely impacts the small-scale galaxy cross-correlation function. Choice of AGN feedback prescription in Horizon-AGN could play a role in driving the discrepancy with VIDEO data – to test this we would measure the clustering in the twin simulation without AGN feedback.

Finally, we will look towards using analyses like that presented in this text to understand how to best make inferences from non-linear galaxy clustering in the future using *Euclid* and LSST. The huge amounts of data collected in these surveys will mean that systematic uncertainties will dominate over statistical uncertainties and understanding observational biases like those explored here will be essential for making the most of these planned facilities in the coming years.

## ACKNOWLEDGEMENTS

Many thanks to the anonymous referee whose comments have greatly improved the quality of this paper. PWH wishes to acknowledge support provided through an STFC studentship. Many thanks to Jeremy Blaizot for comments and input to the ideas in this paper and to Steven Murray for advice for using HALOMOD. CL was supported by a Beecroft Fellowship. JD acknowledges funding support from Adrian Beecroft, the Oxford Martin School, and the STFC. OI acknowledges the funding of the French Agence Nationale de la Recherche for the project ‘SAGACE’. We thank Stephane Rouberol for running the cluster hosted by the Institut d’Astrophysique de Paris. This work was supported by the Oxford Centre for Astrophysical Surveys, which is funded through generous support from the Hintze Family Charitable Foundation, the award of the STFC consolidated grant (ST/N000919/1), and the John Fell Oxford University Press (OUP) Research Fund. This research was also supported in part by the National Science Foundation under Grant No. NSF PHY-1748958. Based on data products from observations made with ESO Telescopes at the La Silla or Paranal Observatories under ESO programme ID 179.A-2006. Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada–France–Hawaii Telescope, which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l’Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada–France–Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

## REFERENCES

- Agertz O., Kravtsov A. V., 2015, *ApJ*, 804, 18
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, 310, 540
- Arnouts S. et al., 2002, *MNRAS*, 329, 355
- Artale M. C. et al., 2017, *MNRAS*, 470, 1771
- Aubert D., Pichon C., Colombi S., 2004, *MNRAS*, 352, 376

- Baldry I. K. et al., 2010, *MNRAS*, 404, 86
- Beckmann R. S. et al., 2017, *MNRAS*, 472, 949
- Beltz-Mohrmann G. D., Berlind A. A., Szewciw A. O., 2019, preprint ([arXiv:1908.11448](https://arxiv.org/abs/1908.11448))
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, *MNRAS*, 360, 159
- Blanton M., Cen R., Ostriker J. P., Strauss M. A., 1999, *ApJ*, 522, 590
- Blanton M., Cen R., Ostriker J. P., Strauss M. A., Tegmark M., 2000, *ApJ*, 531, 1
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Castelló-Mor N., Netzer H., Kaspi S., 2016, *MNRAS*, 458, 1839
- Cen R., Ostriker J. P., 2000, *ApJ*, 538, 83
- Chabrier G., 2003, *PASP*, 115, 763
- Chisari N. E. et al., 2018, *MNRAS*, 480, 3962
- Chiu I. et al., 2016, *MNRAS*, 458, 379
- Coupon J. et al., 2012, *A&A*, 542, A5
- Coupon J. et al., 2015, *MNRAS*, 449, 1352
- Crocce M. et al., 2016, *MNRAS*, 455, 4301
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Croton D. J. et al., 2016, *ApJS*, 222, 22
- Dalton G. B. et al., 2006, in McLean I. S., Iye M., eds, *Proc. SPIE Conf. Ser. Vol. 6269, Ground-Based and Airborne Instrumentation for Astronomy*. SPIE, Bellingham, p. 62690X
- Davé R., Thompson R., Hopkins P. F., 2016, *MNRAS*, 462, 3265
- Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, *MNRAS*, 420, 2662
- Dubois Y. et al., 2014, *MNRAS*, 444, 1453
- Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, *MNRAS*, 463, 3948
- Durkalec A. et al., 2015, *A&A*, 583, A128
- Farrow D. J. et al., 2015, *MNRAS*, 454, 2120
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Guo H., Zehavi I., Zheng Z., 2012, *ApJ*, 756, 127
- Gwyn S. D. J., 2012, *AJ*, 143, 38
- Hatfield P. W., Jarvis M. J., 2017, *MNRAS*, 472, 3570
- Hatfield P. W., Lindsay S. N., Jarvis M. J., Häußler B., Vaccari M., Verma A., 2016, *MNRAS*, 459, 2618
- Hearin A. P., Behroozi P. S., van den Bosch F. C., 2016, *MNRAS*, 461, 2135
- Ilbert O. et al., 2006, *A&A*, 457, 841
- Jakobs A. et al., 2018, *MNRAS*, 480, 3338
- Jarvis M. J. et al., 2013, *MNRAS*, 428, 1281
- Johnston R., Vaccari M., Jarvis M., Smith M., Giovannoli E., Häußler B., Prescott M., 2015, *MNRAS*, 453, 2541
- Katsianis A., Tescari E., Wyithe J. S. B., 2016, *PASA*, 33, e029
- Kaviraj S. et al., 2017, *MNRAS*, 467, 4739
- Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, *MNRAS*, 450, 1349
- Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, *MNRAS*, 451, 2900
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Laigle C. et al., 2019, *MNRAS*, 486, 5104
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Li C. et al., 2012, *MNRAS*, 419, 1557
- Limber D. N., 1954, *ApJ*, 119, 655
- McAlpine K., Smith D. J. B., Jarvis M. J., Bonfield D. G., Fleuren S., 2012, *MNRAS*, 423, 132
- McAlpine S. et al., 2016, *Astron. Comput.*, 15, 72
- McCracken H. J. et al., 2007, *ApJS*, 172, 314
- McCracken H. J. et al., 2015, *MNRAS*, 449, 901
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Neistein E., Weinmann S. M., 2010, *MNRAS*, 405, 2717
- Pawlowski M. S., Ibata R. A., Bullock J. S., 2017, *ApJ*, 850, 132
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ
- Peirani S. et al., 2017, *MNRAS*, 472, 2153
- Pichon C., Thiébaud E., Prunet S., Benabed K., Colombi S., Sousbie T., Teyssier R., 2010, *MNRAS*, 401, 705
- Pontzen A., Slosar A., Roth N., Peiris H. V., 2016, *Phys. Rev. D*, 93, 103519
- Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, *MNRAS*, 338, 14
- Pozzetti L. et al., 2010, *A&A*, 523, A13
- Pujol A. et al., 2017, *MNRAS*, 469, 749
- Saghiha H., Simon P., Schneider P., Hilbert S., 2017, *A&A*, 601, A98
- Sawala T. et al., 2015, *MNRAS*, 448, 2941
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
- Springel V. et al., 2005, *Nature*, 435, 629
- Springel V. et al., 2018, *MNRAS*, 475, 676
- Stringer M. J., Benson A. J., Bundy K., Ellis R. S., Quetin E. L., 2009, *MNRAS*, 393, 1127
- Teyssier R., 2002, *A&A*, 385, 337
- Trenti M., Stiavelli M., 2008, *ApJ*, 676, 767
- Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, *A&A*, 506, 647
- Vogelsberger M. et al., 2014, *Nature*, 509, 177
- Wechsler R. H., Tinker J. L., 2018, *ARA&A*, 56, 435
- Weinberg D. H., Dave R., Katz N., Hernquist L., 2004, *ApJ*, 601, 1
- Weingartner J. C., Draine B. T., 2001, *ApJ*, 548, 296
- White S. V., Jarvis M. J., Haussler B., Maddox N., 2015, *MNRAS*, 448, 2665
- Yoshikawa K., Taruya A., Jing Y. P., Suto Y., 2001, *ApJ*, 558, 520
- Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, *MNRAS*, 443, 3044
- Zheng Z. et al., 2005, *ApJ*, 633, 791

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.